

# Chapter 1

## Προεπεξεγασία και εξερεύνηση δεδομένων

### 1.1 Εισαγωγή

Σε αυτό το κεφάλαιο περιγράφουμε το πρόγραμμα λήψης δεδομένων και τον προεπεξεργαστή, που αναλαμβάνουν αντίστοιχα τη μαζική λήψη ιστορικών τιμών μετοχών και τη στήσιμο ενός καθαρού, ομογενοποιημένου συνόλου δεδομένων. Ακόμη, αιτιολογούμε την επιλογή των επτά μετοχών ώστε να καλύψουμε διαφορετικά φαινόμενα της αγοράς. Τέλος, κάνουμε μια απλή διερευνητική ανάλυση για να δείξουμε βασικά χαρακτηριστικά των δεδομένων.

### 1.2 Λήψη δεδομένων και προεπεξεργασία

Το πρόγραμμα λήψης δεδομένων είναι ένα μικρό πρόγραμμα Python σχεδιασμένο να απλοποιεί την απόκτηση ιστορικών χρονοσειρών τιμών μετοχών από την υπηρεσία Yahoo. Οι χρήστες παρέχουν μια περιγραφή των δεδομένων σε μορφή `.json` που καθορίζει τα σύμβολα (`tickers`) των μετοχών, τα χρονικά διαστήματα και τη συχνότητα. Στη συνέχεια, το πρόγραμμα ανακτά τα πεδία τιμών. Τα δεδομένα συλλεγονται σε `DataFrames`, ένα για κάθε `ticker`.

Μόλις ληφθούν τα ακατέργαστα δεδομένα, το πρόγραμμα εναρμονίζει ασυνεπείς ονοματολογικές συμβάσεις και δομές σε μια ενιαία διάταξη έξι στηλών. Αυτό το σταθερό σχήμα διασφαλίζει ότι τα εργαλεία προεπεξεργασίας μπορούν να λειτουργήσουν χωρίς προβλή-

ματα. Κάθε χρονοσειρά γράφεται στο δίσκο ως ένα αρχείο .csv: δημιουργούνται φάκελοι κατ' εντολή του χρήστη, οι αριθμητικές τιμές αποδίδονται με σταθερή ακρίβεια, και οι ημερομηνίες ακολουθούν το πρότυπο ISO.

Αφού ολοκληρωθούν όλες οι λήψεις, τα ακατέργαστα δεδομένα προωθούνται στον προεξεργαστή για να μετατραπούν στην τελική τους μορφή, η οποία θα χρησιμοποιηθεί στα πειράματα. Ο προεπεξεργαστής εκτελεί τέσσερεις διαδικασίες:

1. Φόρτωση δεδομένων & βασική τακτοποίηση.
2. Προσαρμογή τιμών.
3. Δημιουργία ομοιόμορφου ημερολογίου.
4. Υπολογισμός λογαριθμικών αποδόσεων 1 ημέρας.

Πρώτα, διαβάζει τα αρχεία και μετατρέπει τη στήλη date σε τύπο `datetime64 (YYYY-MM-DD)`, διότι οι συναρτήσεις ταξινόμησης, ταξινομούν από τα αριστερά προς τα δεξιά. Έπειτα την ορίζει ως δείκτη (`dataFrame index`), για να αξιοποιήσει τις δυνατότητες που παρέχουν οι δείκτες της `pandas`. Παρότι η `Yahoo` παρέχει υψηλής ποιότητας δεδομένα, πρέπει να είμαστε βέβαιοι ότι ο δείκτης έχει αυστηρή μονοτονία (συνεχώς αυξανόμενος χωρίς διπλότυπα). Για να εξασφαλιστεί η μονότονία αυτή, εκτελείται ταξινόμηση βάσει του δείκτη και διαγράφονται τυχόν διπλές καταχωρήσεις.

Οι δημόσια καταχωρημένες εταιρείες, ιδίως οι μεγάλες που μελετώνται εδώ, συχνά πραγματοποιούν διασπάσεις μετοχών. Αυτό σημαίνει ότι αυξάνουν τον απόλυτο αριθμό των μετοχών τους χωρίς να αλλάζει η συνολική αξία της εταιρείας. Τέτοιες διασπάσεις δημιουργούν τεχνητά άλματα τιμών που παραπλανούν τα μοντέλα αλλά και τα μετρικά σφάλματος. Για παράδειγμα, μια διάσπαση 4-σε-1 οδηγεί σε πτώση 75% στην ακατέργαστη χρονοσειρά. Ένα προβλεπτικό μοντέλο μπορεί να το ερμηνεύσει ως κραχ στην αγορά με αποτέλεσμα τα μετρικά σφάλματος να αυξηθούν απότομα και το μοντέλο να μην γενικεύσει σωστά. Για να το αντιμετωπίσει αυτό, ο προεπεξεργαστής χρησιμοποιεί την προσαρμοσμένη τιμή κλεισίματος (`AdjClose`) που παρέχει η `Yahoo`, και υπολογίζει έναν παράγοντα με τον οποίο πολλάπλασιάζονται όλες οι στήλες:

$$\text{factor} = \frac{\text{AdjClose}}{\text{Close}},$$

$$\text{AdjOpen} = \text{factor} \times \text{Open}, \dots$$

Τα ακατέργαστα δεδομένα μπορεί να περιέχουν κενά για ορισμένες ημερομηνίες, τα οποία θα προκαλέσουν ασυμφωνίες κατά την εκπαίδευση μοντέλων. Για να το λύσει αυτό, ο προεπεξεργαστής δημιουργεί ένα ομοιόμορφο ημερολόγιο συναλλαγών από 2010-01-01 έως 2024-12-31. Έπειτα επαναδεικτοδοτεί (`re-index`) το `dataframe` στο νέο δείκτη, ώστε όλα τα σύνολα δεδομένων να έχουν ακριβώς τις ίδιες ημερομηνίες. Οι γραμμές για τις οποίες δεν υπήρχαν τιμές μετατρέπονται σε `NaN`. Επειδή μεγάλα τμήματα συνεχόμενων κενών είναι σπάνια, αρκεί να εφαρμοστεί ένα *forward fill* (όταν βρεις `NaN`, πάρε την τελευταία μη-`NaN` τιμή και αντέγραψε την) και στη συνέχεια *backward fill* (αν το σύνολο ξεκινάει με `NaN` το *forward fill* δεν καλύψει τις αρχικές απούσες τιμές), χώρις φόβο να αλλάξει η κατανομή των τιμών. Τέλος, δημιουργεί μια καινούρια στήλη στην οποία υπολογίζει λογαριθμικές αποδόσεις κάθε ημέρας. Αυτή η στήλη χρησιμοποιείται για τον υπολογισμό τεχνικών δεικτών.

## 1.3 Επισκόπηση συνόλου δεδομένων

Το σύνολο δεδομένων της μελέτης περιλαμβάνει επτά χρονοσειρές ημερήσιων τιμών μετοχών για την περίοδο 2010–2024. Τα ticker που επιλέχθηκαν είναι: `AAPL`, `TSLA`, `XOM`, `SPY`, `JNJ`, `AMD` και `PG`. Κάθε χρονοσειρά περιλαμβάνει τις στήλες: `date`, `open`, `high`, `low`, `close`, `adj_close` και `volume`.

Table 1.1: Παράδειγμα δεδομένων από τη χρονοσειρά `AAPL`

Ημερομηνία	Open	High	Low	Close	Volume
2010-01-04	6.42	6.46	6.39	6.44	493,729,600
2010-01-05	6.46	6.49	6.42	6.45	601,904,800
2010-01-06	6.45	6.48	6.34	6.35	552,160,000
2010-01-07	6.37	6.38	6.29	6.34	477,131,200
2010-01-08	6.33	6.38	6.29	6.38	447,610,800
2010-01-11	6.40	6.41	6.27	6.32	462,229,600

### 1.3.1 Επιλογή Δεδομένων Μετοχών

Η επιλογή των μετοχών σε αυτή τη μελέτη σχεδιάστηκε ώστε να διασφαλίζει μια ολοκληρωμένη αναπαράσταση του τοπίου της χρηματοοικονομικής αγοράς. Επιδιώκει να ισορροπή-

σει την ποικιλία κλάδων, τα καθεστώτα μεταβλητότητας και τους ρόλους τους στη δομή της αγοράς — διαστάσεις που επηρεάζουν σημαντικά τη συμπεριφορά των χρονοσειρών και, κατ' επέκταση, την απόδοση των μοντέλων. Οι επιλεγμένες μετοχές καλύπτουν ένα ευρή φάσμα οικονομικών ρόλων, επιτρέποντας πιο γενικεύσιμα και ερμηνεύσιμα συμπεράσματα μοντελοποίησης.

Οι AAPL και TSLA επιλέχθηκαν για να εκπροσωπήσουν το τμήμα τεχνολογίας υψηλής ανάπτυξης, αλλά καταλαμβάνουν διαφορετικές θέσεις εντός αυτού. Η AAPL (Apple Inc.) είναι μια ώριμη, με παγκόσμια επιρροή εταιρεία, με σταθερά κέρδη, ευρύ οικοσύστημα προϊόντων και μακροχρόνια παρουσία στον δείκτη S&P 500. Η συμπεριφορά της χαρακτηρίζεται από μέτρια μεταβλητότητα και συνεχή ανάπτυξη, καθιστώντας την μετοχή της ως χαμηλού κινδύνου. Αντιθέτως, η TSLA (Tesla Inc.) είναι ένα χαρακτηριστικό παράδειγμα μετοχής υψηλού ρίσκου(high beta growth). Εμφανίζει διακυμάνσεις μεγαλύτερες από την αγορά ως απάντηση σε οικονομικά νέα ή μεταβολές στη διάθεση των επενδυτών. Αυτό οφείλεται στην υποθετική της αξία, την ταχεία επέκταση και την ευαισθησία σε καινοτομία των ανταγωνιστών.

Οι XOM, JNJ και PG επιλέχθηκαν για να καλύψουν κυκλικές και αμυντικές μετοχές. Η XOM (Exxon Mobil) δραστηριοποιείται στον ενεργειακό τομέα, όπου τα έσοδα και οι τιμές συνδέονται στενά με τις διεθνείς αγορές εμπορευμάτων και γεωπολιτικές εξελίξεις· είναι κυκλική μετοχή (στενά συνδεδεμένη με την μακροοικονομική πορεία της οικονομίας), που αποδίδει καλά σε περιόδους οικονομικής ανάπτυξης και υποαποδίδει σε υφέσεις. Αντίθετα, οι JNJ (Johnson & Johnson) και PG (Procter & Gamble) θεωρούνται αμυντικές μετοχές (παρέχουν βασικά αγαθά και υπηρεσίες με σταθερά έσοδα ανεξάρτητα από τις ευρύτερες συνθήκες). Η συμπερίληψη και των τριών μας επιτρέπει να ελέγξουμε πόσο καλά γενικεύουν τα μοντέλα σε κλάδους με διαφορετική έκθεση σε μακροοικονομικές δυνάμεις.

Η SPY είναι ένα ETF που παρακολουθεί την απόδοση του S&P 500, ενός δείκτη 500 μεγάλων εταιρειών των ΗΠΑ. Η SPY χρησιμοποιείται ευρέως στην ακαδημαϊκή χρηματοοικονομική και στην πρακτική διαχείρηση κεφαλαίου, διότι αντικατοπτρίζει τη συνολική κίνηση της αγοράς. Η επιλογή της μας επιτρέπει να αξιολογήσουμε την απόδοση των μοντέλων στο σύνολο της αγοράς.

Τέλος, η AMD επιλέχθηκε για να εκπροσωπήσει τον κλάδο των ημιαγωγών, γνωστός για την κυκλικότητα και την υψηλή μεταβλητότητά του. Η τιμή της μετοχής της διακυμαίνεται σημαντικά λόγω της γρήγορης εξέλιξης της τεχνολογίας των ημιαγωγών, προβλημάτων

στην εφοδιαστική αλυσίδα και έντονου ανταγωνισμού. Αυτά τα χαρακτηριστικά την καθιστούν ιδανική για έλεγχο αντοχής των προβλεπτικών μοντέλων: τα θέτουν αντιμέτωπα με μη-στασιμότητα και έντονα δομικά διαλείμματα(structural breaks). Με την προσθήκη της AMD, αξιολογούμε την ικανότητα ενός μοντέλου να προσαρμόζεται σε αιχμηρές μεταβολές και σύνθετες συμπεριφορές τιμών που σπάνια συναντώνται σε πιο σταθερούς κλάδους.

Συνολικά, ο συνδυασμός αυτός διασφαλίζει ότι τα πειράματά μας θα καλύψουν ένα ευρύ και ρεαλιστικό φάσμα χρηματοοικονομικών συνθηκών, επιτρέποντάς μας να διερευνήσουμε τα δυνατά σημεία και τους περιορισμούς των μοντέλων σε πλαίσια σταθερότητας vs μεταβλητότητας, κυκλικότητας vs αμυντικότητας και επιπέδου εταιρίας vs επιπέδου αγοράς.



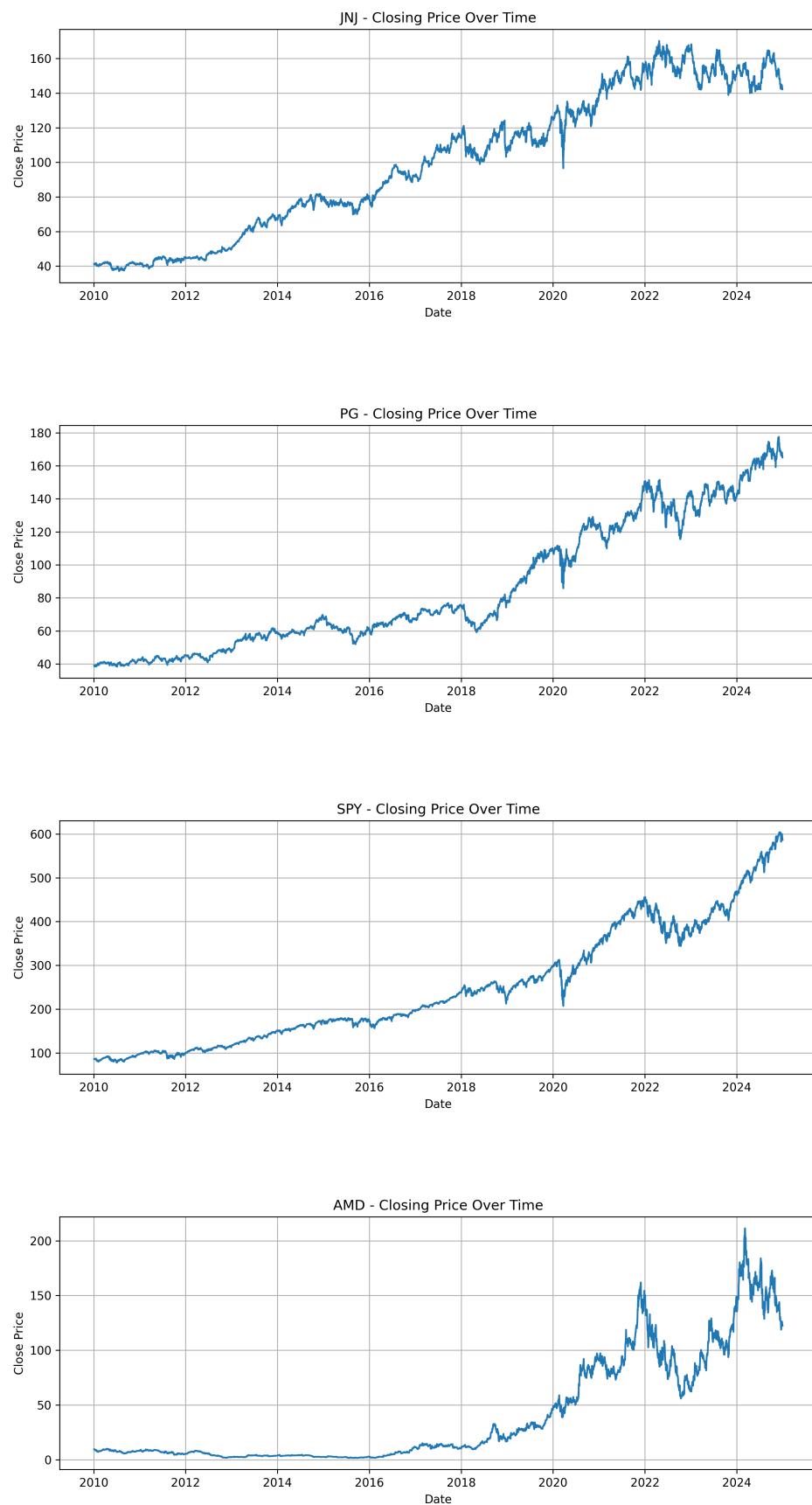


Figure 1.1: Closing Prices

### 1.3.2 Περίοδοι σημαντικών γεγονότων

Από κάθε σύνολο δεδομένων παίρνουμε ένα υποσύνολο, το οποίο αναδεικνύει ενδιαφέρουσες περιπτώσεις που δοκιμάζουν διαφορετικές πτυχές της γενίκευσης των μοντέλων. Ακολουθεί η περιγραφή κάθε γεγονότος, η χρηματοοικονομική του σημασία και ο τρόπος με τον οποίο δοκιμάζει τα προβλεπτικά συστήματα.

**AAPL – Κατάρρευση & Ανάκαμψη (2020-03-10):** Αυτό το παράθυρο αποτυπώνει την βίαιη αντίδραση της αγοράς στην εμφάνιση της πανδημίας COVID-19. Η AAPL γνώρισε απότομη πτώση στις αρχές Μαρτίου 2020 λόγω του παγκόσμιου πανικού και του lockdown. Ακολούθησε ταχεία ανάκαμψη λόγω στροφής των επενδυτών στην τεχνολογία. Αυτό το σχήμα-V δοκιμάζει την ικανότητα ενός μοντέλου να ανταποκρίνεται σε ταχείες αλλαγές μεταβλητότητας.

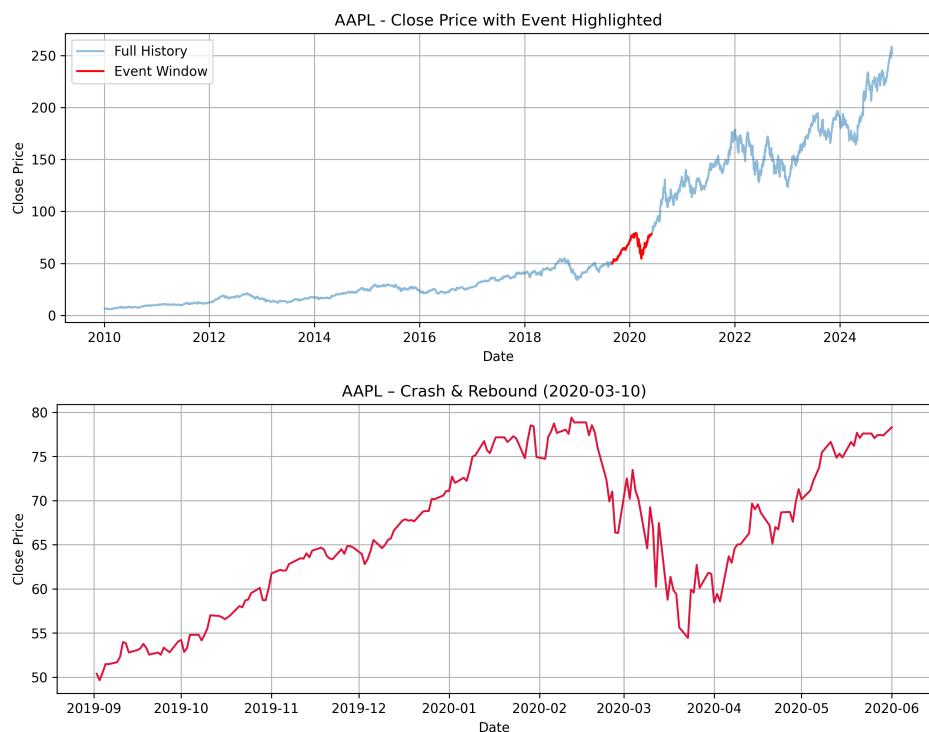


Figure 1.2: AAPL – Κατάρρευση & Ανάκαμψη

**TSLA – Μετοχή Υψηλού Beta (2021-01-15):** Μετά από παραβολική άνοδο το 2020, η TSLA εισήλθε στις αρχές του 2021 σε φάση υψηλής αλλά μη κατευθυνόμενης μεταβλητότητας. Η τιμή διακυμαινόταν σε μεγάλο εύρος. Αυτό το φαινόμενο δείχνει τι συμβαίνει όταν μια μετοχή υψηλού beta χάνει ορμή και οι επενδυτές απομακρύνονται από τα κερδοσκοπικά ονόματα.

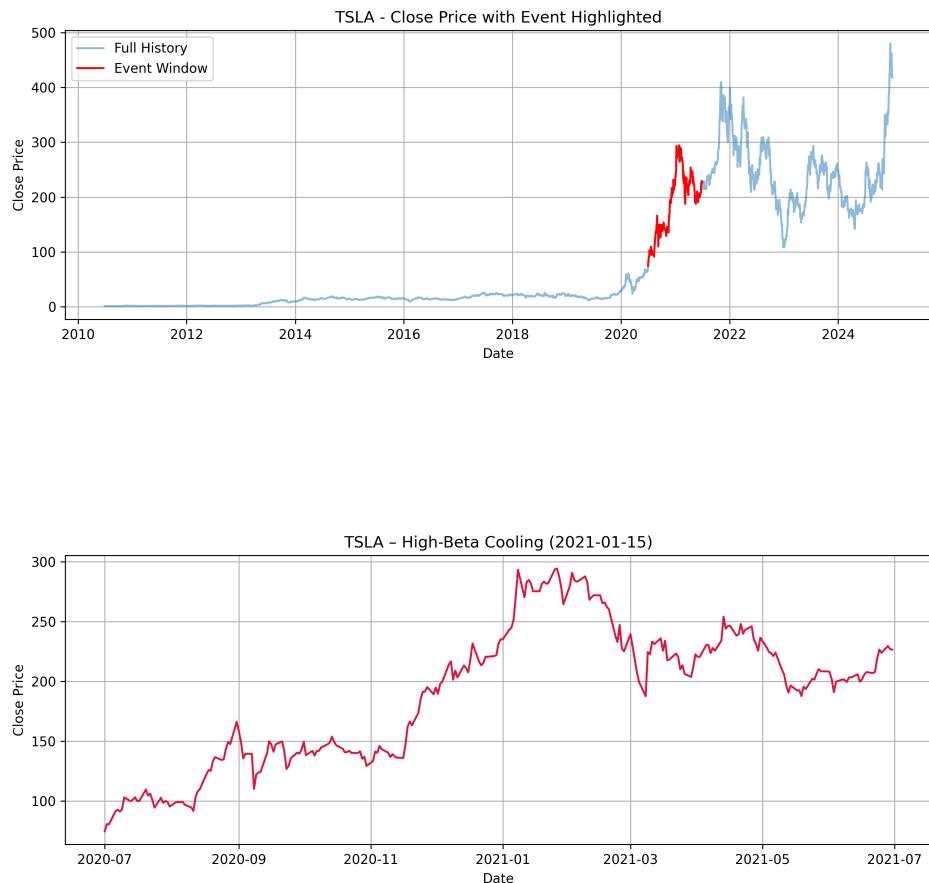
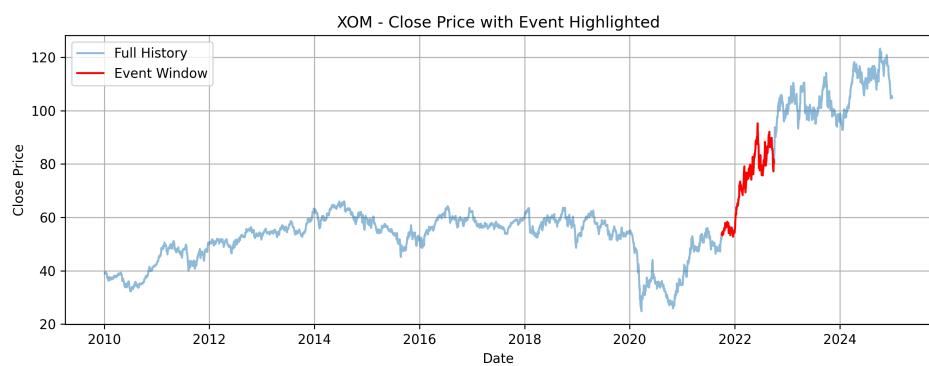


Figure 1.3: TSLA – Μετοχή Υψηλού Beta

**XOM – Κορυφή Πετρελαϊκού Κύκλου (2022-06-01):** Αναπαριστά την κορύφωση του ράλι πετρελαίου μετά την πανδημία, όταν οι τιμές εκτοξεύτηκαν λόγω γεωπολιτικών εντάσεων (π.χ. πόλεμος στην Ουκρανία) και στη συνέχεια κανονικοποιήθηκαν. Η μετοχή της XOM ακολούθησε αυτό το μοτίβο. Η μοντελοποίησης δυσκολεύονται από εξαρτήσεις σε εξωγενείς παράγοντες και από απότομες μεταβάσεις, για αυτό το υποσύνολο είναι ιδανικό τεστ ανίχνευσης αλλαγής κατάστασης στην αγορά.



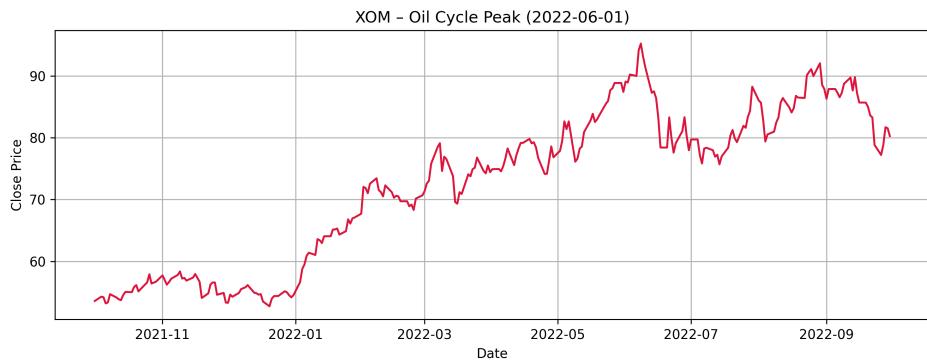


Figure 1.4: XOM – Κορυφή Πετρελαιϊκού Κύκλου

**SPY – Πτώση λόγω μακροοικονομικών παραγόντων (2022-09-15):** Μέσα στο 2022, ο S&P500 μπήκε σε παρατεταμένη φάση πτώσης χωρίς σαφή κατεύθυνση. Σε αντίθεση με ξαφνικά κραχ ή ράλι, εδώ επικρατούσε χαμηλή ορμή και μακροοικονομική αβεβαιότητα (π.χ. πληθωριστικές πιέσεις, αυξήσεις επιτοκίων), δημιουργώντας θορυβώδη κίνηση με ημερήσιες αναστροφές. Τέτοιες περίοδοι είναι δύσκολο να προβλεφθούν, καθώς ο λόγος σήματος θορύβου είναι εξαιρετικά χαμηλός.



Figure 1.5: SPY – Πτώση λόγω μακροοικονομικών παραγόντων

**JNJ – Ανθεκτικότητα (2019-08-01):** Αυτή η φάση δείχνει τη JNJ να κινείται σε στενό

εύρος τιμών, με ελάχιστη μεταβλητότητα για μεγάλο διάστημα. Οι αμυντικές μετοχές, όπως η JNJ, παραμένουν σταθερές σε περιόδους αβεβαιότητας χάρη στον ρόλο τους στην υγεία και τη χαμηλή ζήτηση. Το φαίνομενο αυτό δοκιμάζει τα μοντέλα όχι με πολυπλοκότητα αλλά με στασιμότητα — πώς αποφεύγουν να «φανταστούν» σήματα εκεί όπου δεν υπάρχουν.

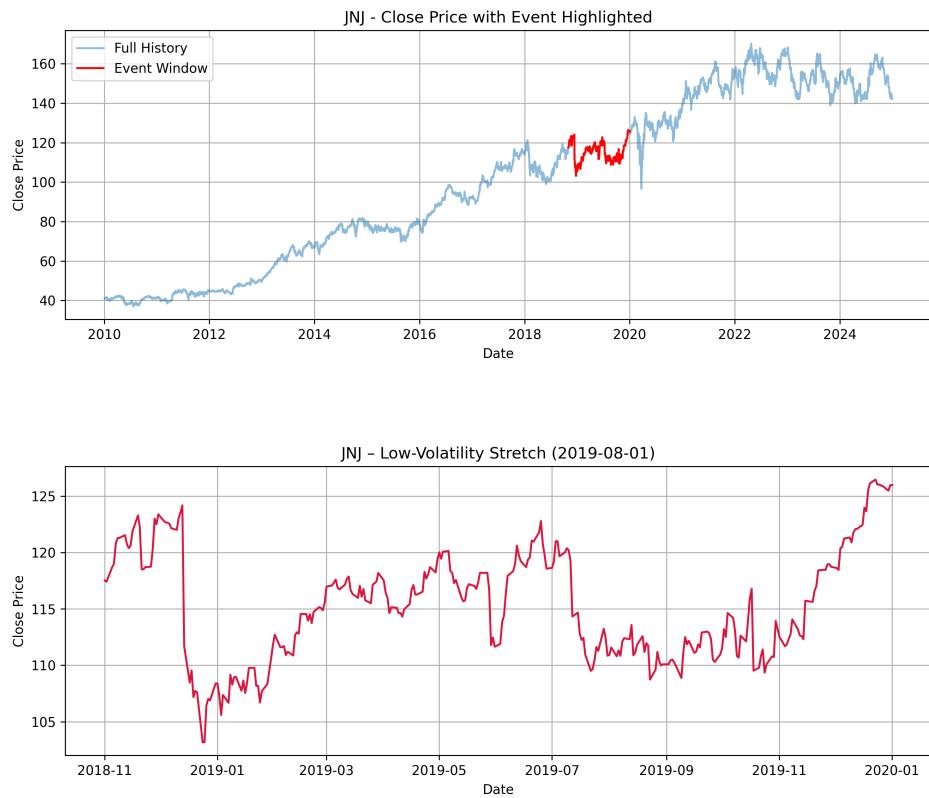


Figure 1.6: JNJ – Ανθεκτικότητα

**AMD – Τεχνολογικό Selloff (2018-10-10):** Περίοδος ευρείας υποτίμησης των τεχνολογικών μετοχών υπό την πίεση ανησυχιών για επιτοκιακές αυξήσεις και γενικό κλίμα κινδύνου. Η τιμή της AMD κατέρρευσε απότομα, απεικονίζοντας συνδεδεμένη πώληση σε high-beta τεχνολογία. Η φάση αυτή δοκιμάζει την ικανότητα ανίχνευσης τομεακών σοκ και γρήγορων μεταβάσεων συσχετίσεων.



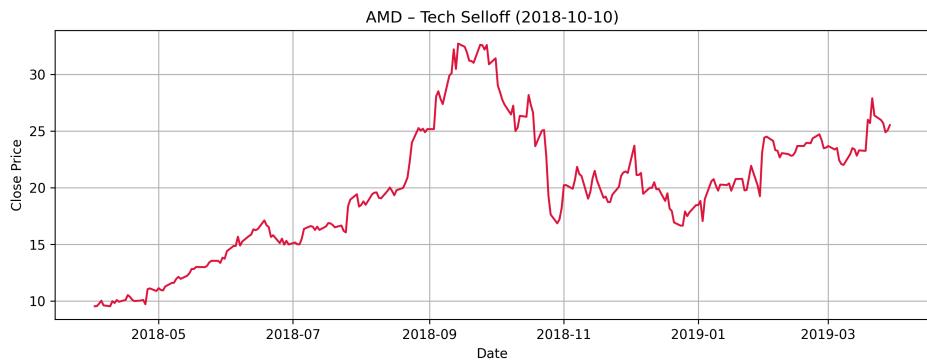


Figure 1.7: AMD – Τεχνολογικό Selloff

**PG – Αγορά Χωρίς Μακροοικονομική Επίδραση (2015-06-15):** Παρουσιάζει τη συμπεριφορά μιας μετοχής καταναλωτικών αγαθών σε περίοδο εξαιρετικά χαμηλής μεταβλητότητας και θορύβου. Η PG κινήθηκε σε στενό εύρος για εβδομάδες ολόκληρες χωρίς να ανταποκριθεί σε ευρύτερα γεγονότα. Τέτοια παράθυρα είναι κρίσιμα για την αξιολόγηση της συντηρητικότητας του μοντέλου.



Figure 1.8: PG – Αγορά Χωρίς Μακροοικονομική Επίδραση

## 1.4 Εξερεύνηση δεδομένων

### 1.4.1 Κινούμενες στατιστικές

Για την εκτίμηση της χρονικής μεταβλητότητας στην απόδοση και τον κίνδυνο, υπολογίζουμε δύο βασικά κινούμενα μεγέθη: τον κινούμενο μέσο των ημερήσιων λογαριθμικών αποδόσεων 21 ημερών και την κινούμενη τυπική απόκλιση αυτών των αποδόσεων 21 ημερών. Ο κινούμενος μέσος  $\mu_t^{(21)}$  απεικονίζει τη τοπική κατεύθυνση της ορμής, ενώ η κινούμενη τυπική απόκλιση  $\sigma_t^{(21)}$  αντανακλά την πρόσφατη μεταβλητότητα:

$$\mu_t^{(21)} = \frac{1}{21} \sum_{i=0}^{20} \log\left(\frac{P_{t-i}}{P_{t-i-1}}\right), \quad \sigma_t^{(21)} = \sqrt{\frac{1}{21} \sum_{i=0}^{20} \left(\log\frac{P_{t-i}}{P_{t-i-1}} - \mu_t^{(21)}\right)^2}.$$

Η AAPL παρουσιάζει μέτρια και σχετικά σταθερή μεταβλητότητα καθ' όλο το δείγμα, με έντονη κορύφωση κατά το κραχ του Μαρτίου 2020 λόγω COVID-19. Οι κινούμενες αποδόσεις της παραμένουν στενά γύρω από το μηδέν, σύμφωνα με την κατηγοριοποίησή της ως ώριμη μετοχή χαμηλού beta.

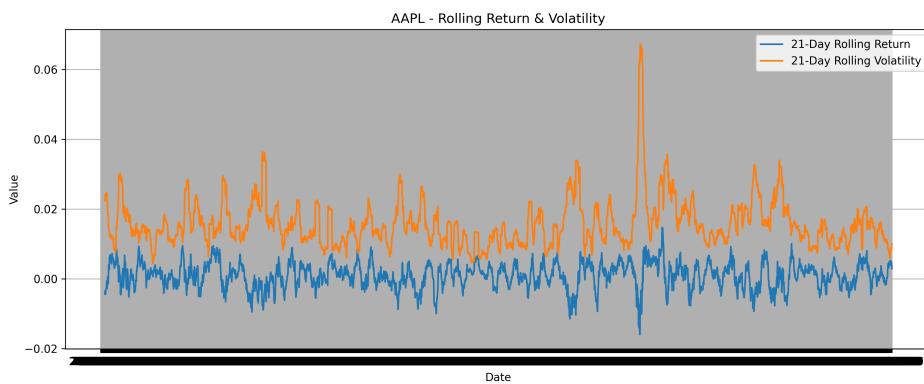


Figure 1.9: AAPL- Κινούμενες στατιστικές

Αντίθετα, η AMD εμφανίζει συχνές και έντονες εκρήξεις μεταβλητότητας, ιδιαίτερα γύρω στο 2018–2019, οι οποίες συμπίπτουν με εφαρμογές κανονισμών στον τεχνολογικό κλάδο, υπογραμμίζοντας τον υψηλό beta χαρακτήρα της. Οι κινούμενες αποδόσεις της είναι επίσης πιο θορυβώδεις, αντανακλώντας την ευαισθησία της στην διάθεση των επενδυτών και σε περιόδους μη -κερδοφορίας.

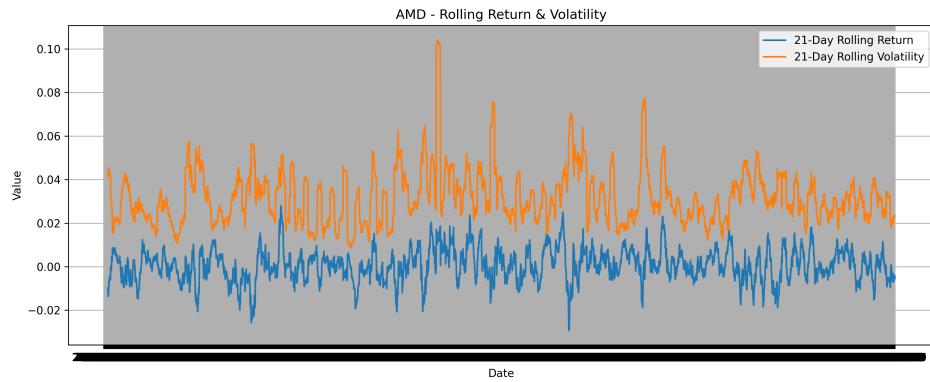


Figure 1.10: AMD- Κινούμενες στατιστικές

Οι JNJ και PG, αμυντικές μετοχές από τους κλάδους υγείας και καταναλωτικών αγαθών, παρουσιάζουν εξαιρετικά σταθερή συμπεριφορά: η μεταβλητότητά τους σπάνια υπερβαίνει το 0.015 και οι αποδόσεις τους παραμένουν στενά συγκεντρωμένες γύρω από τον μέσο όρο. Αυτές οι δύο χρονοσειρές λειτουργούν ως βάσεις χαμηλής διακύμανσης στο σύνολό μας, χρήσιμα για να διορθώνει μοντέλα που εισάγουν ψευδείς διακυμάνσεις σε ήρεμες περιόδους.

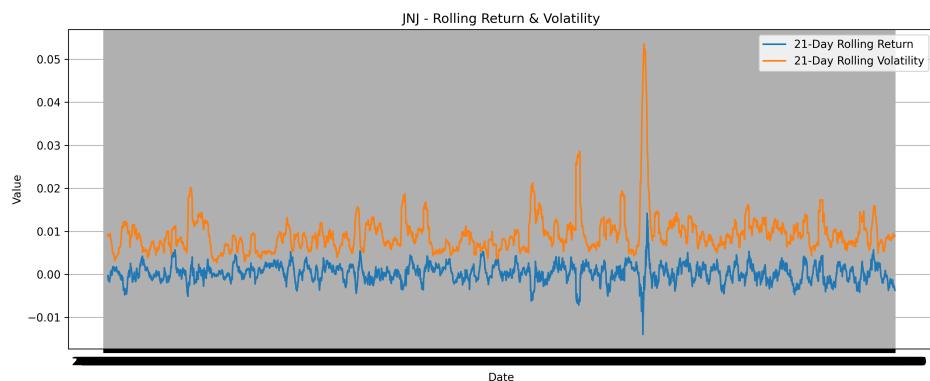


Figure 1.11: JNJ- Κινούμενες στατιστικές

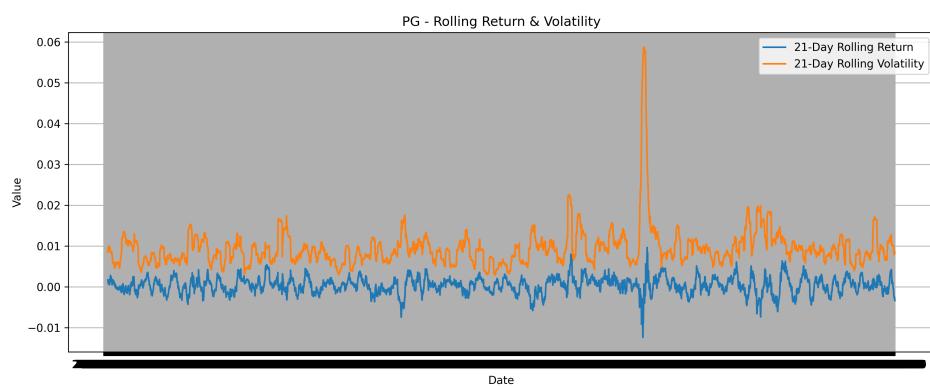


Figure 1.12: PG- Κινούμενες στατιστικές

Η SPY, που παρακολουθεί τον δείκτη S&P 500, συμπεριφέρεται όπως αναμένεται: εμφανίζει έντονη αύξηση της μεταβλητότητας το 2020, αντιστοίχως με τον πανικό της πανδημίας, και αργή επιστροφή σε προ-κρίσης επίπεδα. Οι κινούμενες αποδόσεις της παραμένουν στενά κεντραρισμένες, επιβεβαιώνοντας τον ρόλο της ως σημείο αναφοράς σε επίπεδο αγοράς.

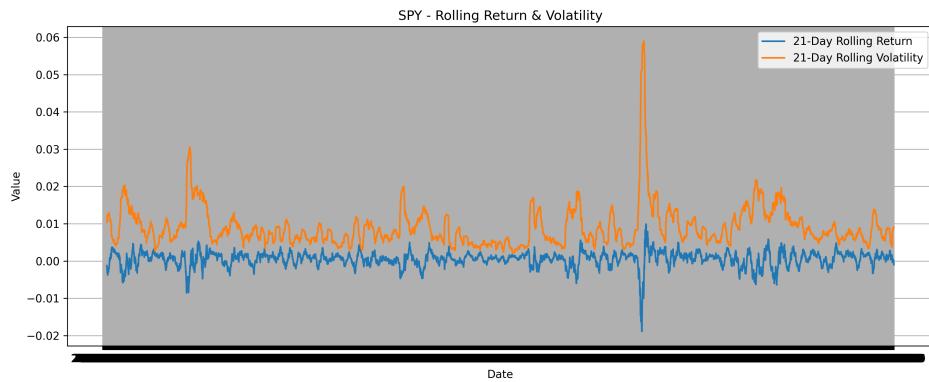


Figure 1.13: SPY - Κινούμενες στατιστικές

Η TSLA παρουσιάζει εντελώς διαφορετικό προφίλ, με επίμονη υψηλή μεταβλητότητα και ακραίες κορυφώσεις, κυρίως κατά το κερδοσκοπικό κύμα του 2020 και τις μετέπειτα περιόδους. Οι κινούμενες αποδόσεις της διακυμαίνονται πολύ περισσότερο σε σχέση με οποιοδήποτε άλλη μετοχή της μελέτης, κάνοντάς τη σημαντική πρόκληση για τα μοντέλα πρόβλεψης.

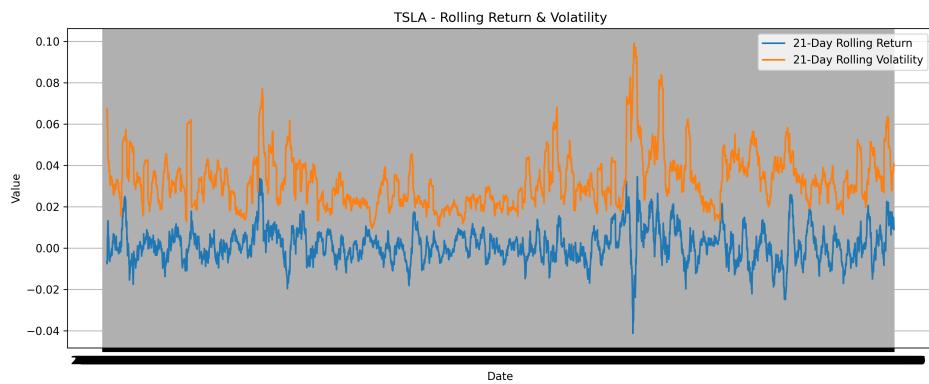


Figure 1.14: TSLA - Κινούμενες στατιστικές

Τέλος, η XOM εμφανίζει ενδιάμεση συμπεριφορά. Η μεταβλητότητά της είναι γενικά μέτρια, αλλά αυξάνεται απότομα κατά το ενεργειακό ράλι του 2022, υπογραμμίζοντας τον αντίκτυπο των μακροοικονομικών γεγονότων στον ενεργειακό τομέα. Οι κινούμενες αποδόσεις της αποκλίνουν επίσης σημαντικά από το μηδέν ανά περιόδους.

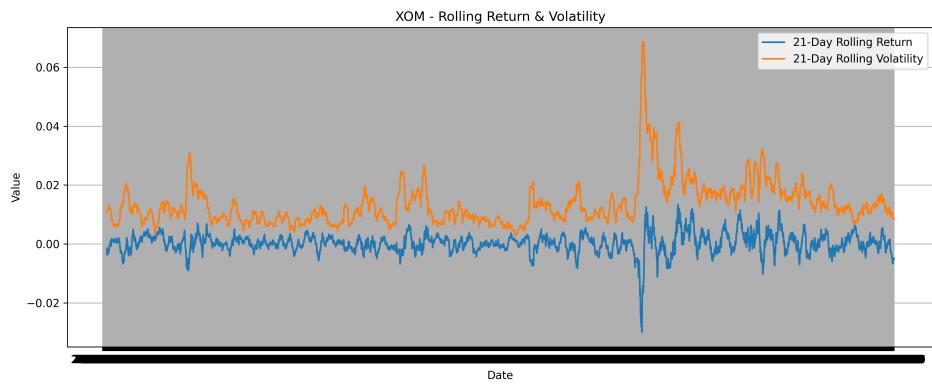


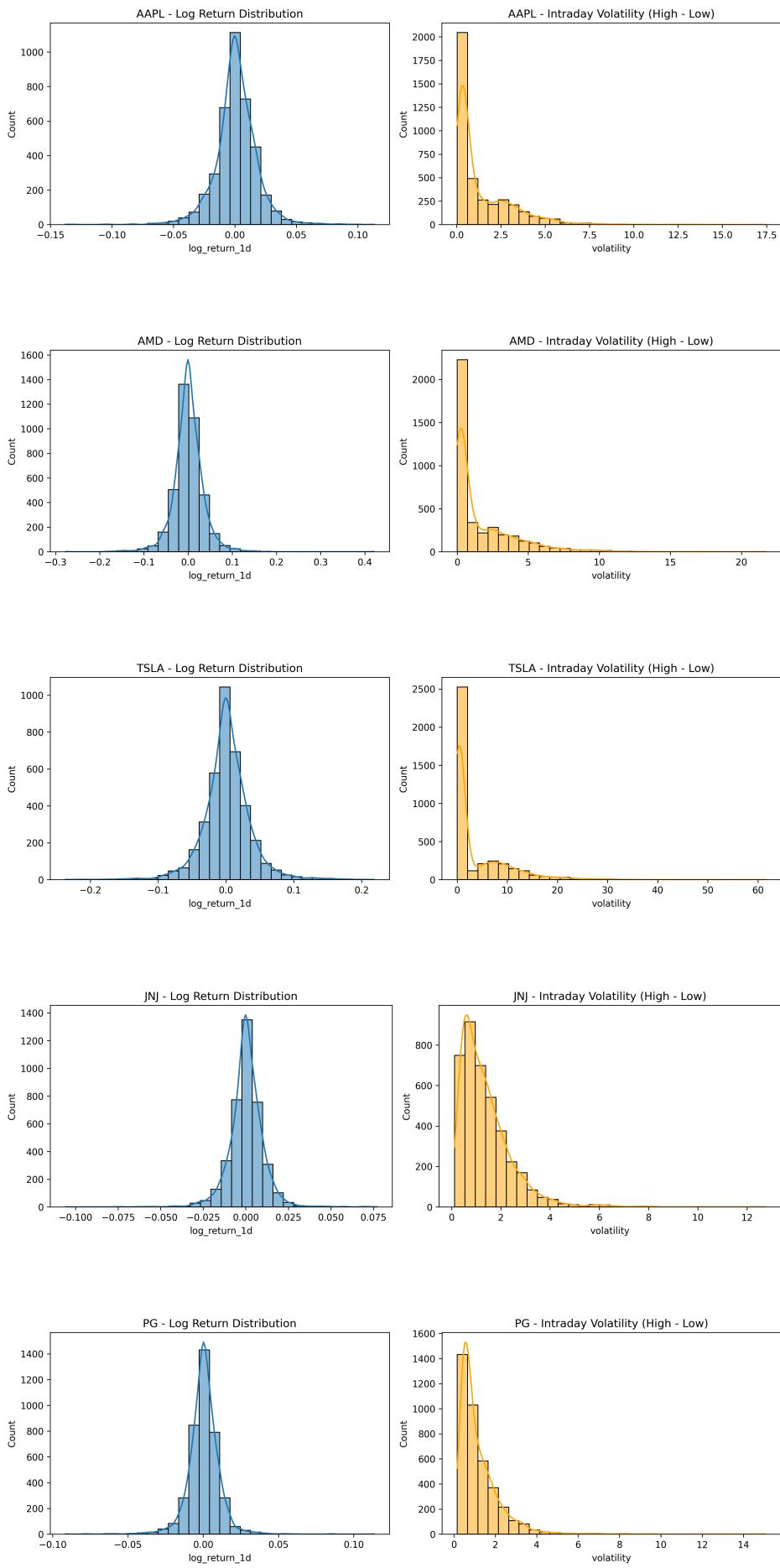
Figure 1.15: XOM - Κινούμενες στατιστικές

## 1.4.2 Χαρακτηριστικά Κατανομών

### Κατανομές ημερήσιων λογαριθμικών αποδόσεων και μεταβλητότητας

Για την καλύτερη κατανόηση των στατιστικών ιδιοτήτων των χρονοσειρών, αναλύσαμε δύο βασικά χαρακτηριστικά κατανομής: την κατανομή των ημερήσιων λογαριθμικών αποδόσεων και την κατανομή της ενδοημερήσιας μεταβλητότητας, ορισμένης ως το ημερήσιο εύρος μεταξύ της υψηλής και της χαμηλής τιμής. Αυτά τα ιστογράμματα, επικαλυμμένα με την συνάρτηση πυκνότητας, αποκαλύπτουν σημαντικές πληροφορίες για την διασπορά τιμών από ημέρα σε ημέρα.

Σε όλα τα στοιχεία, η κατανομή των 1-ημερών λογαριθμικών αποδόσεων είναι κεντραρισμένη γύρω από το μηδέν, αλλά εμφανίζει διαφορετικά επίπεδα κυρτότητας. Η κατανομή της AAPL είναι σχετικά στενή, συμμετρική και μοιάζει με κανονική κατανομή, υποδεικνύοντας χαμηλή μεταβλητότητα και μικρό αριθμό ακραίων τιμών — συνηθισμένα φαινόμενα ώριμων μετοχών. Αντίθετα, το ιστόγραμμα των αποδόσεων της AMD παρουσιάζει σημαντική κυρτότητα, δείχνοντας την παρουσία ακραίων τιμών και αιφνιδιαστικών εκτινάξεων, συνηθισμένων σε μετοχές υψηλού ρίσκου. Ομοίως, η κατανομή της TSLA έχει μεγαλύτερο εύρος τιμών και είναι πιο λεπτοκυρτή σε σχέση με τον μέσο όρο, με βαριές ουρές και ελαφρά ασυμμετρία — χαρακτηριστικά κερδοσκοπικών μετοχών όπου οι κύκλοι ειδήσεων προκαλούν μεγάλες διακυμάνσεις τιμών.



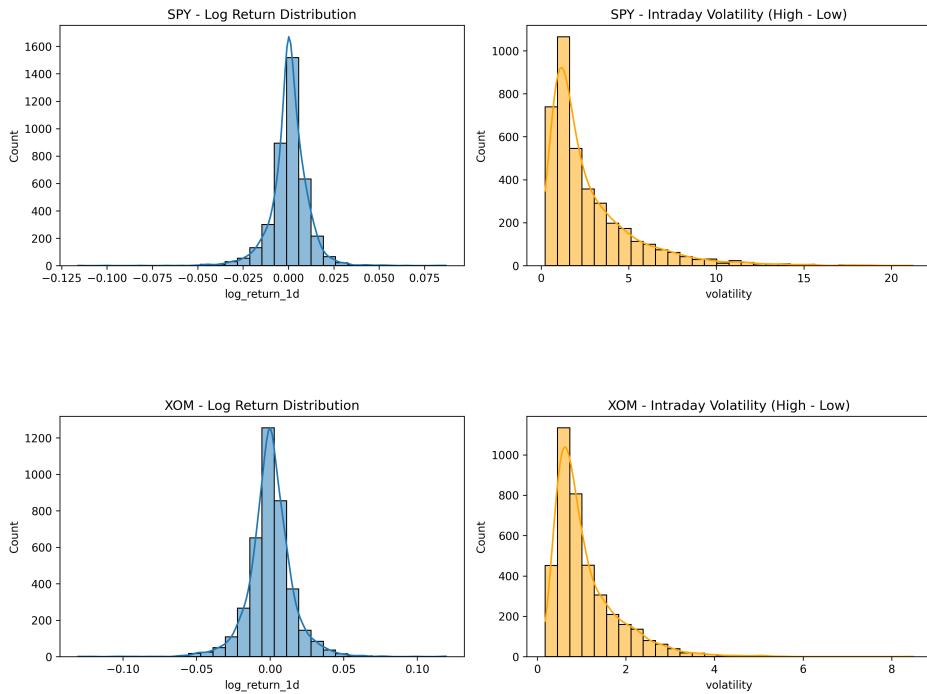


Figure 1.16: Κατανομές log-return/intraday volatility

Οι αμυντικές μετοχές όπως η JNJ και η PG έχουν στενότερες κατανομές αποδόσεων με ελάχιστη ασυμμετρία και κύρτωση.. Οι καμπύλες τους σε σχήμα «καμπάνα» επιβεβαιώνουν τον ρόλο τους ως σταθερές, χαμηλής μεταβλητότητας μετοχές. Η κατανομή των αποδόσεων της SPY θυμίζει κλιμακούμενη κανονική κατανομή. Η συμμετρία και η περιορισμένη συχνότητα ακραίων τιμών αντικατοπτρίζουν τη συγκεντρωτική συμπεριφορά της αγοράς, καθώς ο S&P 500 αφορά 500 εταιρίες. Το ιστόγραμμα της XOM έχει ελαφρώς μεγαλύτερο εύρος από αυτά των αμυντικών μετοχών, με ήπια θετική ασυμμετρία, πιθανόν λόγω των κυκλικών ράλι εμπορευμάτων στον ενεργειακό τομέα.

Στα διαγράμματα ενδοημερήσιας μεταβλητότητας (high - low) παρατηρούμε κατανομές με μεγάλες ουρές και θετική ασυμμετρία σε όλα τα στοιχεία. Αυτό είναι αναμενόμενο, καθώς η μεταβλητότητα δεν μπορεί να είναι αρνητική και ακραίες τιμές εκτείνουν την ουρά δεξιά. Οι αμυντικές μετοχές PG και JNJ συγκεντρώνονται σε χαμηλές τιμές μεταβλητότητας, σπάνια υπερβαίνοντας τις 2–3 ποσοστιαίες μονάδες.

Αντίθετα, η TSLA και η AMD παρουσίαζουν εξαιρετικά ασύμμετρες κατανομές, με συχνά σκαμπανεβάσματα και μεγάλη ουρά που εκτείνεται πάνω από 10%, ακόμη και 20% σε σπάνιες περιπτώσεις. Αυτά τα στοιχεία είναι πιο επιρρεπή σε μεγάλες μεταβολές μίας ημέρας.

Η SPY και η XOM καταλαμβάνουν ενδιάμεση θέση, παρουσιάζοντας μέτρια μεταβλητότητα με περιστασιακές κορυφώσεις.

## Κατανομές Όγκου Συναλλαγών

Για να αξιολογήσουμε την ρευστότητα και την ένταση των συναλλαγών για κάθε εταιρία/δείκτη, εξετάζουμε την κατανομή του ημερήσιου όγκου συναλλαγών. Ο όγκος αντικατοπτρίζει το πόσο ενεργά διαπραγματεύεται μία μετοχή στην αγορά και αποτελεί βασική μεταβλητή, τόσο στη χρηματοοικονομική μοντελοποίηση όσο και σε μικροοικονομικές μελέτες της αγοράς. Τα στοιχεία με υψηλό και σταθερό όγκο προσφέρουν συνήθως καλύτερο λόγο σήματος–θορύβου και διευκολύνουν την ομαλότερη εκπαίδευση μοντέλων. Σε αντίθεση, οι μετοχές με χαμηλό όγκο μπορεί να παρουσιάζουν ιδιόμορφες διακυμάνσεις και έντονες επιδράσεις λόγω ρευστότητας.

Η AAPL ξεχωρίζει με εξαιρετικά υψηλό και διαρκές ημερήσιο όγκο, συγκεντρωμένο σταθερά στο εύρος των 200–400 εκατομμυρίων μετοχών. Η κατανομή της είναι θετικά ασύμμετρη, με ουρά που φτάνει έως και 1,5 δισ. μετοχές ημερησίως. Τέτοια ρευστότητα οδηγεί σε υψηλή ακρίβεια πρόβλεψης και κάνει την AAPL στάνταρ μετοχή ελέγχου απόδοσης για οποιδήποτε προβλεπτικό μοντέλο.

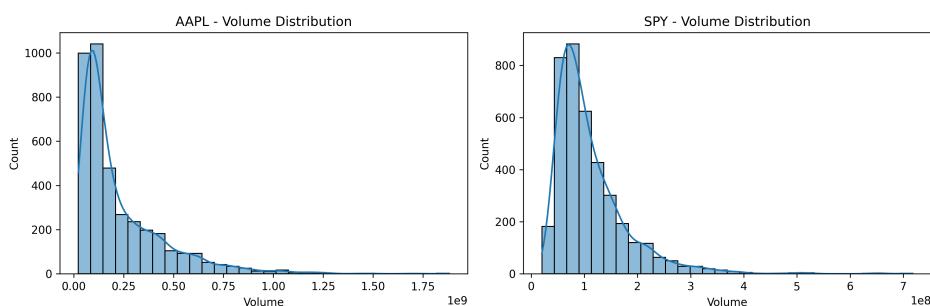


Figure 1.17: Κατανομές όγκου AAPL και SPY

Η SPY παρουσιάζει επίσης σταθερά υψηλό όγκο, κυρίως στο εύρος 50–150 εκατ. μετοχών, με ουρά που εκτείνεται πέραν των 500 εκατ. μετοχών κατά περιόδους έντονης αγοράς. Αυτό δείχνει τον ρόλο της ως κορυφαίο ETF του S&P 500 και συνολικό βαρόμετρο της διάθεσης και των προσδοκιών της αγοράς.

Οι TSLA και AMD εμφανίζουν υψηλή ασυμμετρία, με κύριες τιμές όγκου στις 50–100 εκατ. μετοχές, αλλά συχνές κορυφώσεις κατά τις ημέρες ανακοίνωσης αποτελεσμάτων ή σε περιόδους υψηλής κερδοφορίας. Αυτές οι κατανομές υποδηλώνουν εποχική ρευστότητα που κινείται από ειδήσεις και δραστηριότητα επενδυτών.

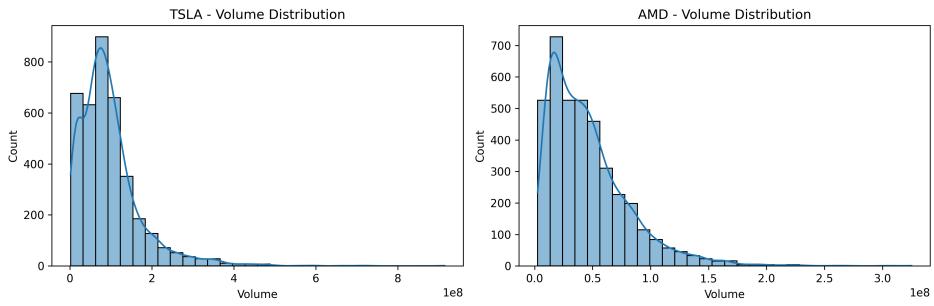


Figure 1.18: Κατανομές όγκου TSLA και AMD

Αντίθετα, οι JNJ και PG παρουσιάζουν συμπαγείς και σταθερές κατανομές όγκου, συγκεντρωμένες γύρω στις 10–30 εκατ. μετοχές. Ο χαμηλός τους όγκος και μεταβλητότητα υποδηλώνει ότι καθοδηγούνται περισσότερο από σταθερότητα και αμυντική κατανομή κεφαλαίων παρά από βραχυπρόθεσμη κερδοσκοπία. Αυτά τα χαρακτηριστικά τις καθιστούν πολύτιμα τεστ για μοντέλα σε ομαλές, χαμηλού θορύβου συνθήκες.

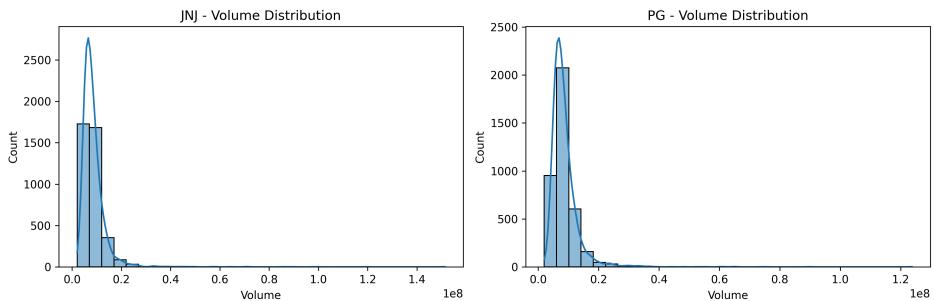


Figure 1.19: Κατανομές όγκου JNJ και PG

Η XOM βρίσκεται στο μέσο του φάσματος όγκου, με κατανομή συγκεντρωμένη γύρω στις 20–60 εκατ. μετοχές. Η ρευστότητά της αντιδρά σε μακροοικονομικά γεγονότα, ιδιαίτερα στις αγορές πετρελαίου και ενέργειας, όπως φαίνεται από περιστασιακά άλματα κατά σημαντικές γεωπολιτικές ή εμπορευματικές εξελίξεις.

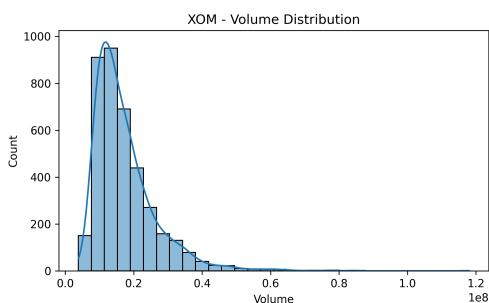


Figure 1.20: Κατανομή όγκου XOM

Συνολικά, η ποικιλία στα χαρακτηριστικά όγκου του δείγματος ενισχύει το πλαίσιο μοντελοποίησής μας, εκθέτοντας το σύστημα τόσο σε εξαιρετικά ρευστές, σταθερές αγορές όσο και σε μεσαίου ύψους χαμηλού όγκου, εποχικές συνθήκες συναλλαγών. Αυτό είναι κρίσιμο για την αξιολόγηση της ανθεκτικότητας των προβλεπτικών μοντέλων υπό διαφορετικά επίπεδα δραστηριότητας της αγοράς.

### 1.4.3 Υπολογισμός Δείκτη Sharpe

Για να ποσοτικοποιήσουμε τη σχέση μεταξύ απόδοσης και κινδύνου, υπολογίζουμε τον δείκτη Sharpe για κάθε μετοχή του συνόλου δεδομένων. Το μέτρο αυτό χρησιμοποιείται ευρέως στη χρηματοοικονομική ανάλυση για να εκφράσει την υπερβάλλουσα απόδοση ανά μονάδα μεταβλητότητας. Υποθέτοντας μηδενικό επιτόκιο για απλότητα, ο δείκτης Sharpe δίνεται από την:

$$\text{Sharpe} = \frac{\mathbb{E}[r]}{\text{Std}(r)} \sqrt{252}$$

Τα αποτελέσματα αποκαλύπτουν σαφείς διαφορές στην απόδοση κάθε μετοχής. Η AAPL ξεχωρίζει με τον υψηλότερο δείκτη Sharpe, ξεπερνώντας το 0.8, κάτι που υποδηλώνει ότι οι αποδόσεις της είναι όχι μόνο συνεπείς αλλά και σχετικά σταθερές — καθιστώντας την αποδοτική στην παραγωγή κερδών για τους επενδυτές. Οι SPY και TSLA ακολουθούν, παρουσιάζοντας ισχυρές απόδοσεις με διαφορετικές όμως μεταβλητότητες: η SPY επωφελείται από την ευρεία ποικιλία, ενώ η υψηλή τιμή του Sharpe για την TSLA προκύπτει από τη διαχρονικά ισχυρή της δυναμική, παρά τα αυξημένα επίπεδα ρίσκον.

Οι PG και JNJ εμφανίζουν μέτριους δείκτες Sharpe γύρω στο 0.5. Αυτό υποδηλώνει ότι, παρότι οι αποδόσεις τους είναι χαμηλότερες σε σχέση με τις μετοχές ανάπτυξης, είναι πιο σταθερές και συνεπώς προσφέρουν αξιοπρεπείς αποδόσεις σε σχέση με τον κίνδυνο. Αντίθετα, οι AMD και XOM παρουσιάζουν τους χαμηλότερους δείκτες Sharpe στο δείγμα. Για την AMD, η χαμηλή τιμή αποδίδεται στη μεγάλη μεταβλητότητα που υπερκαλύπτει τη συνέπεια στις αποδόσεις. Στην περίπτωση της XOM, οι κυκλικές διακυμάνσεις που συνδέονται με τον ενεργειακό κλάδο εισάγουν μακροοικονομική έκθεση και πτωτικές περιόδους, μειώνοντας την αποδοτικότητά της ως προς τον κίνδυνο.

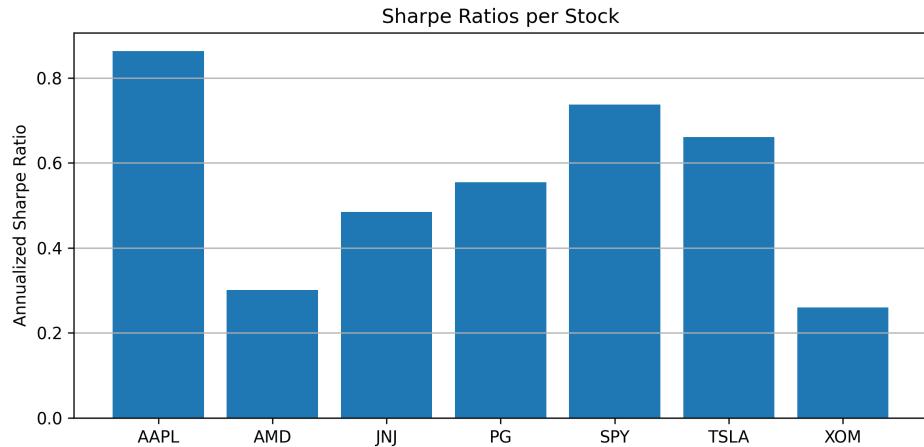


Figure 1.21: Δείκτης sharpe για τις 7 μετοχές

Τα ευρήματα αυτά ενισχύουν τα διαφοροποιημένα προφίλ απόδοσης–κινδύνου του συνόλου δεδομένων μας και υποδεικνύουν ότι τα μοντέλα που αξιολογούνται με αυτά τα δεδομένα πρέπει να χειρίζονται εξίσου αποτελεσματικά τόσο metox;ew υψηλής απόδοσης/υψηλού κινδύνου όσο και χαμηλής μεταβλητότητας/χαμηλής απόδοσης.

#### 1.4.4 Συσχέτιση Ημερήσιων Αποδόσεων

Για να εξετάσουμε τη ταυτόχρονη κίνηση και τον πλεονασμό μεταξύ των μετοχών, υπολογίζουμε τον πίνακα συσχέτισης Pearson των λογαριθμικών ημερήσιων αποδόσεων 1 ημέρας. Ο πίνακας αυτός αποτυπώνει τον βαθμό στον οποίο οι μετοχές κινούνται μαζί σε ημερήσια βάση, προσφέροντας πληροφορίες για τις δυνατότητες διαφοροποίησης και για τις κοινές αντιδράσεις σε καθεστώτα της αγοράς. Τιμές κοντά στο 1 υποδηλώνουν ισχυρή θετική συσχέτιση, ενώ τιμές κοντά στο 0 υποδεικνύουν ανεξαρτησία ή ορθογωνιότητα κινήσεων.

Όπως αναμενόταν, η SPY εμφανίζει τη μεγαλύτερη συνολική συνδεσιμότητα, με μέτριες έως ισχυρές θετικές συσχετίσεις με σχεδόν όλα τα άλλα στοιχεία—ιδιαίτερα με την AAPL (0.69), την XOM (0.61) και την JNJ (0.59). Αυτό αντανακλά τον ρόλο της ως ευρέος δείκτη αγοράς: οι κινήσεις της αποτελούν σύνθεση των μετοχών που περιλαμβάνει, οι οποίες συμπεριλαμβάνουν όλα τα ονόματα στο δείγμα μας.

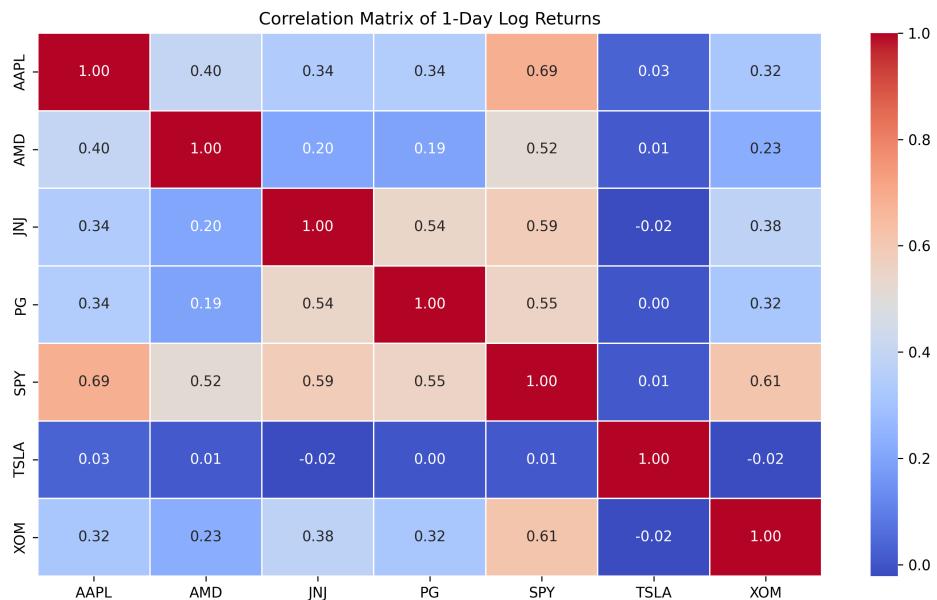


Figure 1.22: Πίνακας Pearson

Η AAPL συσχετίζεται μέτρια με την AMD (0.40), δειχνοντας τη συσχέτιση εντός του τεχνολογικού τομέα. Οι JNJ και PG εμφανίζουν ισχυρό εσωτερικό δεσμό (0.54), σύμφωνο με την κοινή τους κατηγορία ως αμυντικές, χαμηλής μεταβλητότητας μετοχές. Αυτές οι συσχετίσεις είναι σημαντικές για την κατανόηση του πώς οι σταθεροί κλάδοι συμπεριφέρονται από κοινού υπό μακροοικονομικές αλλαγές.

Η TSLA εμφανίζεται σε μεγάλο βαθμό ασυσχέτιστη με τις υπόλοιπες μετοχές. Οι συσχετίσεις της κυμαίνονται κοντά στο μηδέν με κάθε άλλη μετοχή—συμπεριλαμβανομένων αρνητικών ή σχεδόν μηδενικών τιμών με τις JNJ, PG και XOM. Αυτή η αποσυσχέτιση οφείλεται στους ιδιοσυγκρασιακούς παράγοντες της TSLA, όπως κύκλους καινοτομίας, διάθεση λιανικών επενδυτών και συχνές αποκλίσεις από ευρύτερους οικονομικούς δείκτες.

Η AMD, αν και ανήκει στον ίδιο τεχνολογικό τομέα με την AAPL και την TSLA, εμφανίζει όπως είπαμε μόνο ασθενείς συσχετίσεις με αυτές (0.40 με την AAPL, αλλά 0.01 με την TSLA). Αυτό υπογραμμίζει εκ νέου την ετερογένεια της συμπεριφοράς των επενδυτών ακόμη και μεταξύ στενά συνδεδεμένων μετοχών.

Συνολικά, η δομή συσχέτισης επιβεβαιώνει ότι το δείγμα περιέχει ένα ετερογενές μείγμα στοιχείων με κοινά σήματα και ανεξάρτητες, θορυβώδεις δυναμικές. Αυτό εξασφαλίζει ότι τα μοντέλα αξιολογούνται τόσο σε περιβάλλοντα κοινού σήματος (π.χ. AAPL, SPY, AMD) όσο και σε απο-συσχετισμένες, θορυβώδεις περιόδους (π.χ. TSLA).

## 1.5 Συμπέρασμα

Η ανάλυση αποκαλύπτει πλούσια ποικιλία συμπεριφορών σε πλήρη ιστορικά δεδομένα και παράθυρα γεγονότων. Συνολικά, το σύνολο δεδομένων προσφέρει ένα ισορροπημένο και ρεαλιστικό πλαίσιο για την αξιολόγηση προβλεπτικών μοντέλων σε διαφορετικά επίπεδα σταθερότητας, κινδύνου και συσχέτισης.